見出しの探索頻度を考慮した探索路長の考察†

中 村 良 三†† 松 山 公 一†††

見出し探索法を評価する基準のひとつである探索路長は，これまで主として見出しの探索頻度が一様であると仮定したもとで論議されているが，現実の問題では探索頻度は個々に異なるものである．したがって，見出しの探索頻度を考慮した探索路長の評価を行うことができれば，現実に合った探索路長の評価のみならず探索方法の適切な比較や最適な探索技法の構成が可能になると考える．本論文では，見出しの任意の探索頻度に対して，その平均探索路長および分散を評価できる表現式を，逐次探索法，２分木探索法，分散記憶法について，見通しよく，系統的に導き出す．次に，この表現式を用いて，探索頻度に代表的な二，三の確率分布を与えたときの表現式と数値例を示し，具体例を用いて各探索方法における探索路長を比較検討する．

1. まえがき

見出し探索の方法は，一般に探索に要する時間，すなわち探索路長のほか挿入，削除等の操作を含んだ速度および情報を格納する領域以外に操作に要する領域としてのポインタなどをあわせた所要領域の大きさ，すなわち空間効率などによって評価されている．

評価基準のひとつである探索路長は，これまで主として各見出しの探索頻度が一様と仮定したもとで論議されているが，現実の問題では探索頻度は見出し個々に異なるものである．

したがって，見出しの探索頻度を考慮した探索路長の評価を行うことができれば，現実に合った探索路長の評価のみならず探索方法の適切な比較や最適な探索技法の構成が可能になると考える．

一般に，ある見出しを探すために要する見出しの比較回数，すなわち探索する路の長さは，見出しが配置される場所に左右され，その置かれる場所は見出しが登録される順序に依存する．

探索路長の評価に際しては，見出しの探索頻度が一様と仮定したときには，探索路長が見出しの登録順序に依存しないものもあるが，探索頻度を考慮に入れれば，探索路長は各見出しの探索頻度のみならず登録順序に依存するので，これらを把握する必要がある．

本論文では，見出しの探索頻度を考慮し，探索路長の分布特性を把握できる平均探索路長および分散を，基本的な探索方法である逐次探索法，２分木探索法および分散記憶法の内，分離連鎖法と線形法について，評価する表現式を見通しよく，系統的に導き出す．

次に，これらの評価式を用いて，見出しの探索頻度を具体的に与えたときの平均探索路長および分散を数量的に把握し論議する．

2. 導出する表現式の一般概念

見出しの探索頻度を考慮した探索路長を導出するためには，各見出しの探索頻度およびその配置場所を把握する必要がある．

そこで，見通しよく，系統的な定式化を行うため，探索回数を確率変数にとり，その確率分布を考察して，平均探索路長および分散を評価できる表現式を導き出す．

導出の過程で，見出しの個数をN，表の大きさをM，各見出しが探索される確率を登録順序に従い ρ_i $(i=1,2,\cdots,N;\ \sum_{i=1}^{N}\rho_i=1)$ とし，探索が成功するときの平均探索路長および分散を S_N, V_N とし，不成功のときのこれらを $\bar{S}_N, \bar{V}_N$ とする．

探索が不成功のときには，探索路長は探索頻度に依存しないので，本論文では主として探索が成功するときの探索路長を次のような三つの頻度分布について論議する．

a）見出しが探索される確率が一様のとき

このとき，$\rho_i=\dfrac{1}{N}$ $(i=1,2,\cdots,N)$ となる．

b）見出しが探索される確率が登録順序に従い半減するとき

このとき，$\rho_i=\dfrac{C}{2^{i-1}}$ $(i=1,2,\cdots,N)$ となる．

ただし，$C=\dfrac{1}{2-2^{1-N}}$ とする．

† Considerations of the Number of Probes Deliberating the Probability Distribution for the Frequency of Probes by RYOZO NAKAMURA and KIMIKAZU MATSUYAMA (University of Kumamoto).

†† 熊本大学電子計算機室
††† 熊本大学

c）見出しが探索される確率が登録順序に従い調和減少するとき

このとき，$\rho_i=\frac{C}{i}$ $(i=1,2,\cdots,N)$ となる．

ただし，$C=\frac{1}{H_N}$ とする．

ここで，H_N は $1+\frac{1}{2}+\cdots+\frac{1}{N}$ を表す．

3. 逐次探索法

この初歩的な探索法では，見出しを登録するときには表の形で確保されている領域に見出しを含むレコードを登録順にすき間なく表の先頭から格納し，見出しを探索するときには，与えられた見出しを表の先頭から逐次格納されている見出しと照合する方法をとる．

このとき，探索路長の表現式は次のように導き出される．

Ⅰ）成 功 探 索

見出しの比較回数，すなわち探索回数 k を確率変数にとり，探索される確率 ρ_k を考慮すると，平均探索路長および分散は次のように表現される．

$$S_N=\sum_{k=1}^{N}k\rho_k \tag{1}$$

$$V_N=\sum_{k=1}^{N}k^2\rho_k-S_N^2 \tag{2}$$

この表現式を用いて，見出しが探索される確率が一様な場合，登録順序に従い半減する場合および調和減少する場合の探索路長を次に示す．

a）探索される確率が一様なとき

$$S_N=\frac{N+1}{2} \tag{3}$$

$$V_N=\frac{N^2-1}{12} \tag{4}$$

b）探索される確率が登録順序に従い半減するとき

$$S_N=2-\frac{N}{2^N-1} \tag{5}$$

$$V_N=2-\frac{N^2}{2^N-1}-\left(\frac{N}{2^N-1}\right)^2 \tag{6}$$

c）探索される確率が登録順序に従い調和減少するとき

$$S_N=\frac{N}{H_N} \tag{7}$$

$$V_N=\frac{N^2+N}{2H_N}-\left(\frac{N}{H_N}\right)^2 \tag{8}$$

ここで，a)，b)，c）の場合の数値例を**表 1** に示す．

表 1 逐次探索法における成功探索路長の比較
Table 1 Comparisons of the search length in successful searches by sequential search.

Size of identifiers	Uniform		Half		Harmonic	
	S_N	V_N	S_N	V_N	S_N	V_N
10	5.5	8.25	1.990	1.902	3.414	7.121
20	10.5	33.25	1.999	1.999	5.559	27.466
30	15.5	74.91	1.999	1.999	7.509	60.004
40	20.5	133.25	1.999	1.999	9.349	104.250
50	25.5	208.25	1.999	1.999	11.113	159.882

Ⅱ）不成功探索

このときには，表に登録されているすべての見出しを調べることになるから探索路長は次のようになる．

$$\bar{S}_N=N+1 \tag{9}$$

$$\bar{V}_N=0 \tag{10}$$

4. 2分木探索法

この探索法では，無作為に生成された見出しを逐次2分木に登録していくとき，n 個の見出しによって生成されている2分木に $n+1$ 番目の見出しを登録するとき k 回の比較を要するような見出しの順列の数を C_{nk} とすれば，C_{nk} は再帰的に次のように表現できる．

$$C_{nk}=2C_{n-1,k-1}+(n-1)C_{n-1,k} \tag{11}$$

このとき，(11)から生成される数列の母関数 $G(z)$ は次のように定義される．

$$G(z)=(2z+n-1)(2z+n-2)\cdots 2z$$

$$=\sum_{k=0}^{n}\begin{bmatrix}n\\k\end{bmatrix}^{*}2^kz^k \tag{12}$$

(12)は n 個の見出しによって生成されている2分木に $n+1$ 番目の見出しが k 回の比較で挿入される可能な場所が $\begin{bmatrix}n\\k\end{bmatrix}2^k$ 個あると解釈される．

また，任意の比較回数で挿入される場所，すなわち $n+1$ 番目の見出しの挿入可能場所は $(n+1)!$ 個ある．したがって，$n+1$ 番目の見出しが k 回の比較で登録される確率を P_{nk} とすれば，P_{nk} は次のように表現される．

$$P_{nk}=\frac{\begin{bmatrix}n\\k\end{bmatrix}2^k}{(n+1)!}(k=0,1,2,\cdots,n) \tag{13}$$

そこで，N 個の見出しから成る2分木を探索するときには登録されたときと同じ探索路を辿るので，(13)の表現式を用いれば，j 回の探索で見つかる確率は $\rho_jP_{j-1,j-1}+\rho_{j+1}P_{j,j-1}+\cdots+\rho_NP_{N-1,j-1}(j=1,2,\cdots,N)$ となる．

* [:] はスターリングの第1種の数を示す[1]．

このことから，探索路長の表現式は次のようになる．

I）成 功 探 索

このとき，平均探索路長および分散は次のように表される．

$$S_N=\sum_{k=0}^{N-1}(k+1)\sum_{j=k}^{N-1}\rho_{j+1}P_{jk}$$
$$=\sum_{k=0}^{N-1}\rho_{k+1}\sum_{j=0}^{k}(j+1)P_{kj} \tag{14}$$
$$V_N=\sum_{k=0}^{N-1}\rho_{k+1}\sum_{j=0}^{k}(j+1)^2P_{kj}-S_N^2 \tag{15}$$

ただし，$P_{k0}=\begin{cases}0 & (k>0)\\ 1 & (k=0)\end{cases}$

(14)，(15) に (13) を代入すれば，(14)，(15) は次のようになる．

$$S_N=\sum_{k=0}^{N-1}\rho_{k+1}\sum_{j=0}^{k}(j+1)\begin{bmatrix}k\\j\end{bmatrix}2^j/(k+1)!$$
$$=\sum_{k=0}^{N-1}\rho_{k+1}\frac{1}{(k+1)!}\left\{\sum_{j=0}^{k}j\begin{bmatrix}k\\j\end{bmatrix}2^j+\sum_{j=0}^{k}\begin{bmatrix}k\\j\end{bmatrix}2^j\right\}^{*}$$
$$=\sum_{k=0}^{N-1}\rho_{k+1}\{2H_{k+1}-1\}$$
$$=2\sum_{k=1}^{N}\rho_kH_k-1 \tag{16}$$
$$V_N=\sum_{k=0}^{N-1}\rho_{k+1}\sum_{j=0}^{k}(j+1)^2\begin{bmatrix}k\\j\end{bmatrix}2^j/(k+1)!-S_N^2$$
$$=\sum_{k=0}^{N-1}\rho_{k+1}\frac{1}{(k+1)!}\left\{\sum_{j=0}^{k}j^2\begin{bmatrix}k\\j\end{bmatrix}2^j\right.$$
$$\left.+2\sum_{j=0}^{k}j\begin{bmatrix}k\\j\end{bmatrix}2^j+\sum_{j=0}^{k}\begin{bmatrix}k\\j\end{bmatrix}2^j\right\}^{**}-S_N^2$$
$$=\sum_{k=0}^{N-1}\rho_{k+1}(4H_{k+1}^2-2H_{k+1}-4H_{k+1}^{(2)}+3)$$
$$-\left(2\sum_{k=0}^{N-1}\rho_{k+1}H_{k+1}-1\right)^2$$
$$=4\sum_{k=1}^{N}\rho_kH_k^2+2\sum_{k=1}^{N}\rho_kH_k-4\sum_{k=1}^{N}\rho_kH_k^{(2)}$$
$$-\left(2\sum_{k=1}^{N}\rho_kH_k\right)^2+2 \tag{17}$$

ただし，$H_N^{(s)}$ は $1+\frac{1}{2^s}+\frac{1}{3^s}+\cdots+\frac{1}{N^s}$ を表す．

(16) の S_N は導出過程は異なるが文献2）の表現式に一致する．

ここで，探索頻度を仮定したときの探索路長を求めると次のようになる．

a）探索頻度が一様なとき

$$S_N=\frac{2}{N}\sum_{k=1}^{N}H_k-1$$
$$=2\left(1+\frac{1}{N}\right)H_N-3 \tag{18}$$
$$V_N=\frac{4}{N}\sum_{k=1}^{N}H_k^2+\frac{2}{N}\sum_{k=1}^{N}H_k$$
$$-\frac{4}{N}\sum_{k=1}^{N}H_k^{(2)}-\frac{2}{N^2}\left(\sum_{k=1}^{N}H_k\right)^2+2$$
$$=\left(2+\frac{10}{N}\right)H_N-4\left(1+\frac{1}{N}\right)\left(H_N^{(2)}+\frac{H_N^2}{N}\right)+4 \tag{19}$$

b）探索頻度が登録順序に従い半減するとき

$$S_N=\frac{2}{1-2^{-N}}\sum_{k=1}^{N}\frac{H_k}{2^k}-1 \tag{20}$$
$$V_N=\frac{2}{1-2^{-N}}\sum_{k=1}^{N}\frac{2H_k^2+H_k-2H_k^{(2)}}{2^k}$$
$$-\left(\frac{2}{1-2^{-N}}\right)^2\left(\sum_{k=1}^{N}\frac{H_k}{2^k}\right)^2+2 \tag{21}$$

c）探索頻度が登録順序に従い調和減少するとき

$$S_N=\frac{2}{H_N}\sum_{k=1}^{N}\frac{H_k}{k}-1$$
$$=H_N+\frac{H_N^{(2)}}{H_N}-1 \tag{22}$$
$$V_N=\frac{2}{H_N}\sum_{k=1}^{N}\frac{2H_k^2+H_k-2H_k^{(2)}}{k}$$
$$-\frac{4}{H_N^2}\left(\sum_{k=1}^{N}\frac{H_k}{k}\right)^2+2 \tag{23}$$

ここで，a)，b)，c）の数値例を**表2**に示す．

II）不成功探索

不成功時の平均探索路長 $\bar{S}_N$ はN個の見出しで生成された2分木を探索するとき，見出しが見つからない場合の探索回数の平均であるが，これは $N+1$ 番目の見出しを登録するときの探索回数の平均と同値である．したがって，$\bar{S}_N, \bar{V}_N$ は次のように表現される．

$$\bar{S}_N=\sum_{k=1}^{N}kP_{Nk}$$

表 2　2分木探索法における成功探索路長の比較
Table 2　Comparisons of the search length in successful searches by binary tree search.

Size of identifiers	Uniform		Half		Harmonic	
	S_N	V_N	S_N	V_N	S_N	V_N
10	3.443	2.193	1.769	0.916	2.458	2.097
20	4.555	3.572	1.772	0.928	3.041	3.647
30	5.256	4.459	1.772	0.928	3.398	4.764
40	5.771	5.107	1.772	0.928	3.657	5.647
50	6.178	5.615	1.772	0.928	3.860	6.383

* $G(z)=\sum_{j=0}^{k}\begin{bmatrix}k\\j\end{bmatrix}2^jz^j$ とおけば { } 内は $G'(1)+G(1)$ となる．

** $G(z)=\sum_{j=0}^{k}\begin{bmatrix}k\\j\end{bmatrix}2^jz^j$ とおけば { } 内は $G''(1)+3G'(1)+G(1)$ となる．

$$=\frac{1}{(N+1)!}\sum_{k=1}^{N}k\begin{bmatrix}N\\k\end{bmatrix}2^k$$
$$=2(H_{N+1}-1) \tag{24}$$
$$\bar{V}_N=\sum_{k=1}^{N}k^2P_{Nk}-\bar{S}_N^2$$
$$=\frac{1}{(N+1)!}\sum_{k=1}^{N}k^2\begin{bmatrix}N\\k\end{bmatrix}2^k-\{2(H_{N+1}-1)\}^2$$
$$=4H_{N+1}^2-6H_{N+1}-4H_{N+1}^{(2)}+6-4(H_{N+1}-1)^2$$
$$=2H_{N+1}-4H_{N+1}^{(2)}+2 \tag{25}$$

5. 分散記憶法

分散記憶法においては衝突の処理の方法によっていろいろな技法が提案されている[2]．

これらの衝突処理技法は同一の分散番地をもつ見出しをひとつのリストに連結する連鎖法と衝突が起こったとき次の登録候補番地を計算によって求める計算法に大別される．

ここでは，連鎖法のなかから分離連鎖法，計算法のなかから線形法を検討の対象に選び，見出しの探索頻度を考慮した探索路長の評価式を導き出す．

このとき，N個の見出しは大きさMの分散表に一様に分散されると仮定する．また，$\alpha\left(=\frac{N}{M}\right)$を表占有率として用いる．

5.1 分離連鎖法

この衝突処理では，同じ分散番地の見出しはその分散番地を探索根とするひとつのリストとして連結する方法をとる．この方法については文献3）で次のように導出している．

N個の見出しを一様にM個のリストに分配するとき，p_{Nh}をリストの長さがhである確率，r_{hj}を長さhのリストの先頭からj番目の見出しが探索される確率，q_{Nk}を任意の長さのリストの先頭からk番目が探索される確率とする．このとき，$q_{Nk}=\sum_{j=k}^{N}r_{jk}p_{Nj}$となる．

このとき，探索路長の表現式は次のようになる．

I）成 功 探 索

このときには探すリストは少なくともひとつ以上の見出しがリンクされているという条件を加味すれば，平均探索路長および分散は次のようになる．

$$S_N=\sum_{k=1}^{N}kq_{Nk}\Big/\sum_{h=1}^{N}p_{Nh} \tag{26}$$
$$V_N=\sum_{k=1}^{N}k^2q_{Nk}\Big/\sum_{h=1}^{N}p_{Nh}-S_N^2 \tag{27}$$

ここで，探索頻度を具体的に考慮すると次のようになる．

a）探索頻度が一様のとき

$$S_N=\frac{1}{2}\left\{\frac{N/M}{1-(1-(1/M))^N}+1\right\}$$
$$\fallingdotseq\frac{1}{2}\left(\frac{\alpha}{1-e^{-\alpha}}+1\right) \tag{28}$$
$$V_N=\frac{(N/M)\cdot(N-1/M)+(N/M)}{3\{1-(1-(1/M))^N\}}-\left\{\frac{N/M}{2(1-(1-(1/M))^N}\right\}^2-\frac{1}{12}$$
$$\fallingdotseq\frac{\alpha(\alpha+1)}{3(1-e^{-\alpha})}-\left\{\frac{\alpha}{2(1-e^{-\alpha})}\right\}^2-\frac{1}{12} \tag{29}$$

b）探索頻度が登録順序に従い半減するとき

$$S_N=\frac{\frac{N}{1-2^{-N}}\sum_{k=1}^{N}k\sum_{i=k}^{N}\frac{1}{i}\left(\frac{1}{M}\right)^i\left(1-\frac{1}{M}\right)^{N-i}\sum_{j=1}^{N}\binom{N-j}{k-1}\binom{j-1}{i-k}\frac{1}{2^j}}{1-\left(1-\frac{1}{M}\right)^N} \tag{30}$$
$$V_N=\frac{\frac{N}{1-2^{-N}}\sum_{k=1}^{N}k^2\sum_{i=k}^{N}\frac{1}{i}\left(\frac{1}{M}\right)^i\left(1-\frac{1}{M}\right)^{N-i}\sum_{j=1}^{N}\binom{N-j}{k-1}\binom{j-1}{i-k}\frac{1}{2^j}}{1-\left(1-\frac{1}{M}\right)^N}-S_N^2 \tag{31}$$

c）探索頻度が登録順序に従い調和減少するとき

$$S_N=\frac{\frac{N}{H_N}\sum_{k=1}^{N}k\sum_{i=k}^{N}\frac{1}{i}\left(\frac{1}{M}\right)^i\left(1-\frac{1}{M}\right)^{N-i}\sum_{j=1}^{N}\binom{N-j}{k-1}\binom{j-1}{i-k}\frac{1}{j}}{1-\left(1-\frac{1}{M}\right)^N} \tag{32}$$
$$V_N=\frac{\frac{N}{H_N}\sum_{k=1}^{N}k^2\sum_{i=k}^{N}\frac{1}{i}\left(\frac{1}{M}\right)^i\left(1-\frac{1}{M}\right)^{N-i}\sum_{j=1}^{N}\binom{N-j}{k-1}\binom{j-1}{i-k}\frac{1}{j}}{1-\left(1-\frac{1}{M}\right)^N}-S_N^2 \tag{33}$$

表 3 分離連鎖法における成功探索路長の比較（分散表の大きさ $M=50$）

Table 3 Comparisons of the search length in successful searches by separate chaining in the scatter storage techniques. (Size of the scatter table $M=50$)

Size of identifiers	Load factor	Uniform		Half		Harmonic	
		S_N	V_N	S_N	V_N	S_N	V_N
25	0.5	1.13	0.14	1.25	0.26	1.20	0.21
50	1.0	1.29	0.34	1.56	0.63	1.45	0.52
100	2.0	1.65	0.89	2.29	1.54	2.06	1.35
150	3.0	2.07	1.63	3.13	2.59	2.78	2.39
200	4.0	2.53	2.55	4.06	3.65	3.56	3.57
250	5.0	3.01	3.64	5.02	4.75	4.38	4.85

ここで，a)，b)，c) の数値例を**表3**に示す．

II) 不成功探索

この場合には，見出しがひとつもないリストを探すかリストの最後尾まで探すかのいずれであるから，平均探索路長および分散は次のようになる．

$$\bar{S}_N=\sum_{k=0}^{N}(k+\delta_k)p_{Nk}=\frac{N}{M}+\left(1-\frac{N}{M}\right)^N\doteqdot\alpha+e^{-\alpha}\tag{34}$$

$$\bar{V}_N=\sum_{k=0}^{N}(k+\delta_k)^2p_{Nk}-\bar{S}_N^2=\frac{N(M-1)}{M^2}+\left(1-\frac{1}{M}\right)^N\left\{1-\frac{2N}{M}-\left(1-\frac{1}{M}\right)^N\right\}\doteqdot\alpha+e^{-\alpha}(1-2\alpha-e^{-\alpha})\tag{35}$$

ただし，$\delta_k=\begin{cases}1 & (k=0)\\ 0 & (k>0)\end{cases}$

5.2 線 形 法

見出しの値をKとするとき，見出しは分散関数$h(K)$によって，0から $M-1$ の分散番地に変換され，見出しの登録は次のように行うものと仮定する．分散番地で指定された表の位置が空であれば，そこに見出しを格納する．空でなければ，すなわち衝突が起こると，次の登録候補番地として $h(K)-1$，そこでまた衝突が起これば，さらに $h(K)-2,\cdots$ というように空な場所があるまで探し，そこに見出しを格納する．このとき，表の始端すなわち0番地の次は表の終端の $M-1$ 番地，次は $M-2$ 番地というように巡回するものとする．

ここで，n 個の見出しが登録されたとき，次のような関数および確率を文献2）に準じて導入する．

$f(M,n)$: 表の 0 番地が空であるような分散番地列の数．

$g(M,n,h)$: 表の 0 番地が空で，1 番地から h 番地まで見出しで占有され，$h+1$ 番地が空であるような分散番地列の数．

P_{nk}: $n+1$ 番目の見出しを登録するとき，$k+1$ 回の探索を要する確率．

この確率を表すために，文献[2]では，ひとつの添字付 P_k によって表現しているが，探索頻度を考慮するときには各見出しを個々に把握する必要があるので，二つの添字付 P_{nk} とした．

これらの関数および確率は次のように表現される．

$$f(M,n)=\left(1-\frac{n}{M}\right)M^n\tag{36}$$

$$g(M,n,h)=\binom{n}{h}(h+1)^{h-1}(M-h-1)^{n-h-1}(M-n-1)\tag{37}$$

$$P_{nk}=M^{-n}\{g(M,n,k)+g(M,n,k+1)+\cdots+g(M,n,n)\}\tag{38}$$

一方，探索するときには，登録されたと同じ探索路を辿るので，k 回の探索で見つかる見出しは登録されるとき k 回の探索で挿入された見出しであるから，k 回の探索で見つかる確率は $\rho_kP_{k-1,k-1}+\rho_{k+1}P_{k,k-1}+\cdots+\rho_NP_{N-1,k-1}$ $(k=1,2,\cdots,N)$ となる．

したがって，探索路長は次のように導出される．

I) 成 功 探 索

成功時の平均探索路長および分散は次のように求められる．

$$S_N=\sum_{k=0}^{N-1}(k+1)\sum_{j=k}^{N-1}\rho_{j+1}P_{jk}=\sum_{k=0}^{N-1}\rho_{k+1}\sum_{j=0}^{k}(j+1)P_{kj}=\sum_{k=0}^{N-1}\rho_{k+1}\frac{1}{M^k}\sum_{j=0}^{k}\binom{j+2}{2}g(M,k,j)=\frac{1}{2}\sum_{k=0}^{N-1}\rho_{k+1}\left\{1+\frac{1}{M^k}s(k,1,M-1)\right\}=\frac{1}{2}\left\{1+\sum_{k=0}^{N-1}\rho_{k+1}R_0(M,k)\right\}\tag{39}$$

$$V_N=\sum_{k=0}^{N-1}(k+1)^2\sum_{j=k}^{N-1}\rho_{j+1}P_{jk}-S_N^2=\sum_{k=0}^{N-1}\rho_{k+1}\sum_{j=0}^{k}(j+1)^2P_{kj}-S_N^2=\sum_{k=0}^{N-1}\rho_{k+1}\frac{1}{M^k}\left\{\frac{1}{3}\sum_{j=0}^{k}(j+1)^3g(M,k,j)+\frac{1}{2}\sum_{j=0}^{k}(j+1)^2g(M,k,j)\right.$$

$$+\frac{1}{6}\sum_{j=0}^{k}(j+1)g(M,k,j)\Big\}-S_N^2$$

$$=\frac{1}{3}\sum_{k=0}^{N-1}\rho_{k+1}\frac{1}{M^k}t(k,1,M-1)$$

$$-\frac{1}{4}\left\{\sum_{k=0}^{N-1}\rho_{k+1}\frac{1}{M^k}s(k,1,M-1)\right\}^2-\frac{1}{12}$$

$$=\frac{1}{3}\sum_{k=0}^{N-1}\rho_{k+1}R_2(M,k)$$

$$-\frac{1}{4}\left\{\sum_{k=0}^{N-1}\rho_{k+1}R_0(M,k)\right\}^2-\frac{1}{12} \tag{40}$$

ここでは,

$$s(n,x,y)=\sum_{j=0}^{n}\binom{n}{j}(x+j)^{j+1}(y-j)^{n-j-1}(y-n) \tag{41}$$

と定義すれば，次の関係が成立する.

$$s(n,x,y)=x(x+y)^n+ns(n-1,x+1,y-1) \tag{42}$$

同様に,

$$t(n,x,y)=\sum_{j=0}^{n}\binom{n}{j}(x+j)^{j+2}(y-j)^{n-j-1}(y-n) \tag{43}$$

と定義すれば，次の関係が成立する.

$$t(n,x,y)=xs(n,x,y)+nt(n-1,x+1,y-1) \tag{44}$$

また,

$$R_r(M,n)=\sum_{j=0}^{n}\binom{r+j}{r}\binom{n}{j}(j+1)!/M^j \tag{45}$$

と定義すれば，次の関係式が成立する.

$$R_0(M,n)=s(n,1,M-1)/M^n \tag{46}$$

$$R_2(M,n)=t(n,1,M-1)/M^n \tag{47}$$

そこで，探索頻度を具体的に与えると，探索路長は次のようになる.

a）探索頻度が一様のとき

$$S_N=\frac{1}{2}\left\{1+\frac{1}{N}\sum_{k=0}^{N-1}R_0(M,k)\right\}$$

$$=\frac{1}{2}\{1+Q_0(M,N-1)\} \tag{48}$$

$$V_N=\frac{1}{3N}\sum_{k=0}^{N-1}R_2(M,k)$$

$$-\frac{1}{4}\left\{\frac{1}{N}\sum_{k=0}^{N-1}R_0(M,k)\right\}^2-\frac{1}{12}$$

$$=\frac{1}{3}Q_2(M,N-1)-\frac{1}{4}Q_0(M,N-1)^2-\frac{1}{12} \tag{49}$$

ここでは,

$$Q_r(M,n)=\sum_{j=0}^{n}\binom{r+j}{j}\frac{n}{M}\cdot\frac{n-1}{M}\cdot\dots\cdot\frac{n-j+1}{M} \tag{50}$$

と定義すれば[2)]，次の関係が成立する.

$$\frac{1}{(1-\alpha)^{r+1}}-\frac{1}{M}\binom{r+2}{2}\frac{\alpha}{(1-\alpha)^{r+3}}$$

$$\le Q_r(M,N)\le\frac{1}{(1-\alpha)^{r+1}}{}^{*} \tag{51}$$

$$\sum_{k=0}^{n-1}R_0(M,k)=nQ_0(M,n-1) \tag{52}$$

$$\sum_{k=0}^{n-1}R_2(M,k)=nQ_2(M,n-1) \tag{53}$$

(48)の S_N の表現式は導出過程は異なるが文献[2)]の表現式に一致する.

b）探索頻度が登録順序に従い半減するとき

$$S_N=\frac{1}{2}\left\{1+\frac{1}{2-2^{1-N}}\sum_{k=0}^{N-1}\frac{1}{2^k}R_0(M,k)\right\} \tag{54}$$

$$V_N=\frac{1}{3(2-2^{1-N})}\sum_{k=0}^{N-1}\frac{R_2(M,k)}{2^k}$$

$$-\frac{1}{4}\left\{\frac{1}{2-2^{1-N}}\sum_{k=0}^{N-1}\frac{R_0(M,k)}{2^k}\right\}^2-\frac{1}{12} \tag{55}$$

c）探索頻度が登録順序に従い調和減少するとき

$$S_N=\frac{1}{2}\left\{1+\frac{1}{H_N}\sum_{k=0}^{N-1}\frac{R_0(M,k)}{k+1}\right\} \tag{56}$$

$$V_N=\frac{1}{3H_N}\sum_{k=0}^{N-1}\frac{R_2(M,k)}{k+1}$$

$$-\frac{1}{4}\left\{\frac{1}{H_N}\sum_{k=0}^{N-1}\frac{R_0(M,k)}{k+1}\right\}^2-\frac{1}{12} \tag{57}$$

ここで，a)，b)，c)の数値例を**表 4** に示す.

II）不成功探索

N個の見出しが登録されているとき，不成功探索の平均探索路長は $N+1$ 番目の見出しを登録するときの

表 4 線形法における成功探索路長の比較（分散表の大きさ $M=50$）

Table 4 Comparisons of the search length in successful searches by linear probing in the scatter storage techniques. (Size of the scatter table $M=50$)

Size of identifiers	Load factor	Uniform		Half		Harmonic	
		S_N	V_N	S_N	V_N	S_N	V_N
10	0.2	1.106	0.135	1.021	0.023	1.055	0.069
20	0.4	1.292	0.553	1.021	0.023	1.128	0.234
30	0.6	1.628	2.020	1.021	0.023	1.241	0.731
40	0.8	2.377	8.990	1.021	0.023	1.459	2.831
50	1.0	4.771	61.176	1.021	0.023	2.074	17.446

* 文献 2) の p. 530 による.

平均探索路長に等しいので，次の $\bar{S}_N, \bar{V}_N$ が成立する．

$$\begin{aligned}\bar{S}_N&=\sum_{k=0}^{N}(k+1)P_{Nk}\\&=M^{-N}\sum_{k=0}^{N}\binom{k+2}{2}g(M,N,k)\\&=\frac{1}{2}\{1+M^{-N}s(N,1,M-1)\}\\&=\frac{1}{2}\{1+R_0(M,N)\}\\&=\frac{1}{2}\{1+Q_1(M,N)\}\end{aligned}\tag{58}$$

$\bar{S}_N$ は文献[2)]の表現に一致し，(51) から，M が大きく α が 1 の近傍でないときには，次の式で近似できる．

$$\bar{S}_N\fallingdotseq\frac{1}{2}\left\{1+\frac{1}{(1-\alpha)^2}\right\}\tag{59}$$

$$\begin{aligned}\bar{V}_N&=\sum_{k=0}^{N}(k+1)^2P_{Nk}-\bar{S}_N^2\\&=M^{-N}\sum_{k=0}^{N}\left\{\frac{(k+1)^3}{3}+\frac{(k+1)^2}{2}\right.\\&\qquad\left.+\frac{(k+1)}{6}\right\}g(M,N,k)-\bar{S}_N^2\\&=\frac{1}{3M^N}t(N,1,M-1)\\&\qquad-\left\{\frac{1}{2M^N}s(N,1,M-1)\right\}^2\\&\qquad-\frac{1}{12}\\&=\frac{1}{3}\{3Q_3(M,N)\\&\qquad-2Q_2(M,N)\}\\&\qquad-\frac{1}{4}\{Q_1(M,N)\}^2-\frac{1}{12}\\&=Q_3(M,N)-\frac{2}{3}Q_2(M,N)\\&\qquad-\frac{1}{4}Q_1(M,N)^2-\frac{1}{12}\end{aligned}\tag{60}$$

(60) は文献[2)] の問題解答欄で示している表現式に一致する．また，この $\bar{V}_N$ はMが大きく，α が 1 の近傍でないとき，次の式で近似できる．

$$\bar{V}_N\fallingdotseq\frac{3}{4(1-\alpha)^4}-\frac{2}{3(1-\alpha)^3}-\frac{1}{12}\tag{61}$$

6. 比 較 検 討

ALGOL 形式の高級言語で記述されるプログラムでは，使用する名前はあらかじめ宣言定義して用いる．

図 1 は高級言語 PL/I で記述されたプログラム内で使用される 25 個の名前を宣言定義された順序に従い左から右へ並べ，これらの名前が使用された頻度を柱状に表したものである[2)]．

ここでは，図 1 の具体例を用い，前述した各探索方法について，探索頻度が一様と仮定した場合と頻度を考慮した場合における成功時の探索路長を比較検討する．

6.1 逐次探索法

名前が宣言定義された順序に表の先頭から登録されているので，名前の使用頻度を一様と仮定したときに

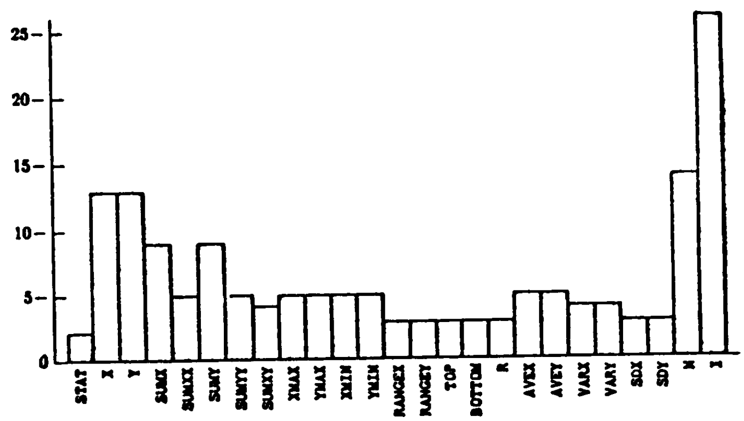

図 1 名前が宣言定義された順序の探索頻度

Fig. 1 The number of times a compiler searches for variable names. The names are listed from left to right in order of their first appearance.

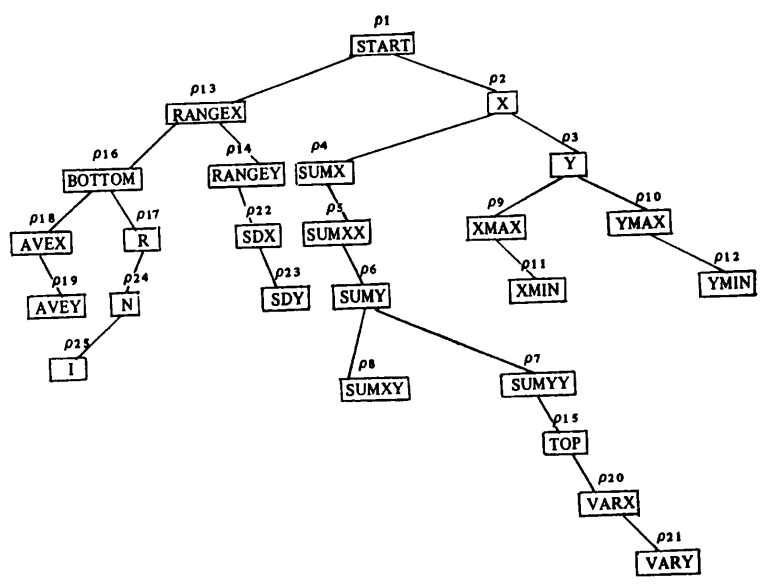

図 2 図 1 の頻度分布を伴った 2 分探索木

Fig. 2 The binary search tree with the frequency distribution on Fig. 1.

は，その平均探索路長は 13.0，その分散は 52.0 であるが，使用頻度を考慮したときには，これらはそれぞれ 13.55，74.59 となる．

使用頻度を考慮したとき，平均，分散ともいくぶん大きくなっているが，それは使用頻度の多い名前（I およびN）が表の後尾に配置されているためと考えられる．

6.2 2分木探索法

名前，すなわち見出しの値を辞書的順序に考えて，名前が宣言定義される順序に従い2分木を生成すると**図2**のようになる．

この2分木で名前の使用頻度を一様と仮定すれば，平均探索路長は 4.52，その分散は 3.29 になるが，頻度を考慮すれば，それらの値は 4.56 および 2.72 となる．

結果的に，平均探索路長には大きな差異は認められないが，分散は頻度を考慮したとき，頻度の多い名前が根の近傍に配置されているため，いくぶん小さくなっている．

一方，見出しが無作為に生成された2分木探索で，見出しが一様な頻度で探索されるような場合は，4章で理論的に導出している表現式(18)，(19)を用いれば，平均探索路長は 4.93，その分散は 4.05 となり，この具体例よりいくぶん悲観的な値になっている．

6.3 分散記憶法

大きさ 50 の分散表に 25 個の名前が一様に分散される場合を想定する．

このとき，名前の使用頻度が一様のときには，分離連鎖法では平均探索路長は 1.13，分散は 0.14 となり，線形法ではそれらの値はそれぞれ 1.43，1.05 となる．

頻度を考慮したときには，分離連鎖法では平均探索路長は 1.14，分散は 0.15 となり，線形法ではそれらの値は 1.50，1.43 となる．

このように，この例では頻度を考慮しても大きな差異は生じない．また，分離連鎖法と線形法でも表の占有率が 0.5 であるので大きな差異は認められない．

この具体例における各探索方法の平均探索路長および分散を**表5**に示す．

表 5 各探索方法における成功探索路長の比較
Table 5 Comparisons of the search length for the respective search methods. (Size of the identifiers $N=25$, Size of the scatter table $M=50$)

Search methods	Uniform probing		Frequency probing	
	S_N	V_N	S_N	V_N
Sequential search	13.00	52.00	13.55	74.59
Binary tree search	4.52	3.29	4.56	2.72
Separate chaining	1.13	0.14	1.14	0.15
Linear probing	1.43	1.05	1.50	1.43

7. むすび

現実に合った探索路長の評価が行えるように，見出しの探索頻度を考慮した平均探索路長および分散を評価する表現式を基本的な探索方法について導き出した．

探索頻度を考慮するためには，探索頻度を一様と仮定するときに比べ，見出し個々の探索頻度のみならず見出しが配置される可能場所を理論的に求め論議する必要がある．

その結果，従来の導出過程に比べ，見通しよく，系統的に式の導出が行えたと考える．

導出した表現式を用いれば，現実に合った探索路長の評価のみならず最適な表の構成や探索方法間の効率の比較が行え，適切な探索方法の選択が可能になると考える．

謝辞 本論文をまとめるにあたり，ご教示いただいた熊本女子大学・数学研究室の城島邦行教授に感謝いたします．

参考文献

1) Knuth, D. E.: *The Art of Computer Programming*, Vol. 1, Addiron-Wesley, Reding (1973).
2) Knuth, D. E.: *The Art of Computer Programming*, Vol. 3, Addison-Wesley, Reading (1973).
3) 中村，松山：分散記憶法における探索頻度を考慮した探索路長とその評価，情報処理学会論文誌，Vol. 24, No. 1, pp. 125-130 (1983).

(昭和57年11月5日受付)
(昭和58年1月17日採録)